JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 75, 447–453 (1980)

A Mixed Theory of Information. V. How to Keep the (Inset) Expert Honest

J. ACZÉL

*Faculty of Mathematics, University of Waterloo,
Waterloo, Ontario N2L 3G1, Canada*

Submitted by R. Bellman

DEDICATED TO THE MEMORY OF C.–F. PICARD

1. This is one in a series of papers (cf.), in particular, Aczél and Daróczy, 1978) on a new, mixed (as distinguished from probabilistic and nonprobabilistic) theory of information.

It will not be supposed, however, that the reader is familiar with the results or even definitions of the previous papers. Also the underlying structure B will be slightly different.

Let B be a Boolean ring of sets (containing the "whole set" Ω and, with any two sets, also their union and difference, thus also their intersection and the empty set 0; see, e.g., Halmos (1950)). We write

$$\Omega_n = \left\{(x_1, x_2, \ldots, x_n) \,\middle|\, \bigcup_{i=1}^{n} x_i = \Omega;\ x_i \in B,\ x_i \cap x_j = 0 \quad \text{if } i \neq j,\quad i,j = 1, 2,\ldots, n\right\}, \tag{1}$$

$$\Omega_n^{\circ} = \left\{(x_1, x_2, \ldots, x_n) \,\middle|\, \bigcup_{i=1}^{n} x_i = \Omega;\ 0 \neq x_i \in B,\ x_i \cap x_j = 0 \quad \text{if } i \neq j,\quad i,j = 1, 2,\ldots, n\right\};$$

$$\Gamma_n = \left\{(p_1, p_2, \ldots, p_n) \,\middle|\, \sum_{i=1}^{n} p_i = 1;\ p_i \geqslant 0,\ i = 1, 2,\ldots, n\right\}, \tag{2}$$

$$\Gamma_n^{\circ} = \left\{(p_1, p_2, \ldots, p_n) \,\middle|\, \sum_{i=1}^{n} p_i = 1;\ p_i > 0,\ i = 1, 2,\ldots, n\right\}.$$

We use *events* x_i as the name for the elements of B, while the p_i are *probabilities.*

The following is a generalization of the problem rhetorically called "how to keep the expert (or forecaster) honest" (see, e.g., McCarthy, 1956; Marschak,

0022-247X/80/060447–07$02.00/0

1959; Good, 1952, 1954; Aczél and Pfanzagl, 1966; Fischer, 1972; Aczél and Ostrowski, 1973; Aczél, 1973, 1974; Aczél and Daróczy, 1975; Walter, 1976). Let the events $x_1, \ldots, x_n$ be results of an experiment (market situation, weather, etc.). We are interested in their probabilities, so we ask an expert. He may know the true probabilities (or, at least, have subjective probabilities) $p_1, p_2, \ldots, p_n$, but tells us $q_1, q_2, \ldots, q_n$ instead. Till now everything is very realistic. Now we make the somewhat idealistic assumption that the expert agrees to be paid, the amount $f_k(x_k, q_k)$, *after* one (and only one) of the events, x_k, happens. So his expected gain is

$$\sum_{k=1}^{n} p_k f_k(x_k, q_k).$$

We want to keep him honest by a method usually applied for the opposite purpose, namely money: we determine the *payoff functions* f_k so that his expected gain is maximal if he told the truth, i.e.,

$$\sum_{k=1}^{n} p_k f_k(x_k, q_k) \leqslant \sum_{k=1}^{n} p_k f_k(x_k, p_k) \tag{3}$$

for all $(x_1, \ldots, x_n) \in \Omega_n$ or $\Omega_n{}^\circ$; $(p_1, \ldots, p_n)$, $(q_1, \ldots, q_n) \in \Gamma_n$ or $\Gamma_n{}^\circ$.

As we see, we allow the payoff functions to depend, in addition to the (true or alleged) probabilities, upon the events themselves or on other parameters associated to the events, for instance, their utility. While no dependence between x_k and p_k, q_k is assumed in general, it seems reasonable to suppose that $x_k = 0$ *implies* $p_k = q_k = 0$. One should require

$$q_k = 0 \Rightarrow p_k = 0, \qquad 0 \cdot f_k(x, 0) = 0 \qquad (k = 1, 2, \ldots, n), \tag{4}$$

since $f_k(x, 0)$ may not be defined [cf. (11), (12), (14), and (15) below]. We will solve the above problem *without any regularity assumption* on f_k. Also, while the problem allows all $n \geqslant 2$ in (3), we will suppose (3) only for one fixed $n > 2$ (the theorems are not true if (3) is supposed only for $n = 2$, see Aczél–Pfanzagl, 1966; Fischer, 1972). However, we emphasize that (3) is supposed for *all* $(x_1, \ldots, x_n) \in \Omega_n$ or $\Omega_n{}^\circ$ and for all $(p_1, \ldots, p_n)$, $(q_1, \ldots, q_n) \in \Gamma_n$ or $\Gamma_n{}^\circ$. The results will turn out to be related to the Shannon entropy and to the inset entropies calculated by Aczél and Daróczy (1978).

2. In the classical, probabilistic theory, the payoff functions f_k do not depend upon the events x_k, only upon the probabilities. Most of the attention has been focused (see the works quoted in the previous paragraph) on the case where all payoff functions are the same $f_1 = f_2 = \cdots = f_n = f$ (cf., however, Good, 1954). For completeness' sake we prove here first a theorem for the

general probabilistic case. (Note: After this paper was finished P. Fischer and P. Kardos informed me that they have found, but not published, the same theorem.)

THEOREM 1. *The inequality*

$$\sum_{k=1}^{n} p_k f_k(q_k) \leqslant \sum_{k=1}^{n} p_k f_k(p_k) \tag{5}$$

is satisfied for one $n > 2$ and for all $(p_1, ..., p_n)$, $(q_1, ..., q_n) \in \Gamma_n{}^\circ$ if, and only if, there exist constants $\alpha \geqslant 0$, $\gamma_1, ..., \gamma_n$ such that

$$f_k(p) = \alpha \log p + \gamma_k \qquad (p \in]0, 1[; k = 1, 2, ..., n). \tag{6}$$

Proof. Choose $p_1 = p$, $q_1 = q$, and $p_i = q_i$ for all $i > 2$. Then [cf. (2)] $p + p_2 = q + q_2 = r$ and (5) reduces to

$$pf_1(q) + (r - p) f_2(r - q) \leqslant pf_1(p) + (r - p) f_2(r - p)$$

or

$$p[f_1(p) - f_1(q)] \geqslant (r - p) [f_2(r - q) - f_2(r - p)] \qquad \text{for all } p, q \in]0, r[, r \in]0, 1[. \tag{7}$$

The domain, on which (7) holds, is symmetric in p and q, so also

$$q[f_1(q) - f_1(p)] \geqslant (r - q) [f_2(r - p) - f_2(r - q)] \tag{8}$$

has to hold on the same domain. Multiplying (7) by $(r - q)$ and (8) by $(r - p)$ and adding the two equations thus obtained, we get

$$r(p - q) [f_1(p) - f_1(q)] \geqslant 0$$

or $p \geqslant q$ implies $f_1(p) \geqslant f_1(q)$; that is, f_1 *is monotonic nondecreasing and, similarly, the same holds for* f_2.

Also from (7) and (8),

$$\frac{f_1(p) - f_1(q)}{p - q}$$

ies between

$$\frac{r - p}{p} \frac{f_2(r - q) - f_2(r - p)}{(r - q) - (r - p)} \quad \text{and} \quad \frac{r - q}{q} \frac{f_2(r - q) - f_2(r - p)}{(r - q) - (r - p)}.$$

Thus, *if* f_2 is *differentiable at* $r - p$, *then* f_1 *is differentiable at* p and

$$pf_1'(p) = (r-p)f_2'(r-p). \tag{9}$$

In other words, *if* f_1 *is not differentiable at* p, *then* f_2 *is not differentiable at any* $r - p \in]0, 1-p[$. But this is impossible, since f_2 is monotonic and thus almost everywhere differentiable. Therefore f_1 (and similarly f_2) *is everywhere differentiable and* (9) *holds for all* $p \in]0, 1[$. So $(s = r - p)$ $pf_1'(p) = sf_2'(s) = \alpha$ (constant) ($\alpha \geqslant 0$, since f is nondecreasing), i.e.,

$$f_1(p) = \alpha \log p + \gamma_1, \qquad f_2(p) = \alpha \log p + \gamma_2$$

and similarly

$$f_k(p) = \alpha \log p + \gamma_k \qquad (p \in]0, 1[;\ \alpha \geqslant 0;\ \gamma_1, \dots, \gamma_n \text{ constants}),$$

which concludes the proof of the "only if" part of Theorem 1. As to the "if" part, it follows immediately from Shannon's inequality (see, e.g., Aczél, 1973; Aczél and Daróczy, 1975)

$$-\sum_{k=1}^{n} p_k \log q_k \geqslant -\sum_{k=1}^{n} p_k \log p_k. \tag{10}$$

The expression on the right side of (10) is *Shannon's entropy*.

3. Now we prove our main results in the "inset" case, that is, where the f_k may also depend upon the events—or subsets—x_k.

THEOREM 2. *Inequality* (3) *holds for a fixed* $n > 2$ *and for all* $(x_1, \dots, x_n) \in \Omega_n$; $(p_1, \dots, p_n), (q_1, \dots, q_n) \in \Gamma_n$ *if, and only if, there exists a constant* $\alpha \geqslant 0$ *and functions* $\gamma_k : B \to R$ $(k = 1, 2, \dots, n)$ *such that* (*with the conventions* $0 \log 0 = 0$ *and* (4))

$$f_k(x_k, 1) \geqslant \gamma_k(x_k), \qquad f_k(x_k, p_k) = \alpha \log p_k + \gamma_k(x_k) \tag{11}$$

$(x_k \in B,\ p_k \in]0, 1[;\ k = 1, 2, \dots, n)$. *Thus the right-hand side of* (3) *reduces to*

$$\alpha \sum_{k=1}^{n} p_k \log p_k + \sum_{k=1}^{n} \gamma_k(x_k)\, p_k \qquad (p_k \in [0, 1];\ k = 1, \dots, n). \tag{12}$$

Proof. Keep, for the time being, x_k constant in (3) and write

$$f_k(p) = f_k(x, p).$$

Then (3) goes over into (5), and we can apply Theorem 1. Taking now the dependence upon x_k into consideration, (6) becomes

$$f_k(x_k, p_k) = \alpha(x_k) \log p_k + \gamma_k(x_k).$$

However, as seen in (6), α is the same for all x_k, so $\alpha(x_1) = \alpha(x_i)$ $(i = 2, 3,..., n)$ and, choosing $x_1 = 0$ [cf. (2), (4)], we see that α *is constant* and we get (11) as asserted. (By (4), $p_k = 1$, $p_j = 0$ $(j \neq k)$ in (3) gives $f_k(x, 1) \geqslant \alpha \log q_k + \gamma_k(x)$. So $f_k(x, 1) \geqslant \gamma_k(x)$.) The "if" part is again a consequence of (10) and of $\alpha \log q \leqslant 0$.

We had to prove that all values of α are the same, not only for sets (elements of B) with empty intersection. This was easy because the substitution of 0 was permissible. However, with a little effort one can prove a similar theorem also if (3) is supposed only for $(x_1 ,..., x_n) \in \Omega_n{}^\circ$; $(p_1, p_2 ,..., p_n)$, $(q_1, q_2 ,..., q_n) \in \Gamma_n{}^\circ$ [cf. (1), (2)].

Indeed, let x_1, y_1 be two arbitrary (nonzero) elements of B, each contained in a $(x_1, x_2 ,..., x_n) \in \Omega_n{}^\circ$, $(y_1, y_2 ,..., y_n) \in \Omega_n{}^\circ$. Since, with x_1, x_2, y_1, y_2, also $(x_1 \backslash y_1) \cup x_2$, $(y_1 \backslash x_1) \cup y_2$, and $x_1 \cap y_1$ are elements of B, so also $(x_1 \cap y_1, (x_1 \backslash y_1) \cup x_2, x_3 ,..., x_n) \in \Omega_n{}^\circ$ and $(x_1 \cap y_1, (y_1 \backslash x_1) \cup y_2, y_3 ,..., y_n) \in \Omega_n{}^\circ$ and thus

$$\alpha(x_1) = \alpha(x_3) = \alpha(x_1 \cap y_1) = \alpha(y_3) = \alpha(y_1).$$

The only case where this does not work, is when $x_1 \cap y_1 = 0$, but then there exists an $i \neq 1$ such that $x_i \cap y_1 \neq 0$ and so $\alpha(x_1) = \alpha(x_i) = \alpha(y_1)$. So the following is proved.

THEOREM 3. *Inequality* (3) *holds for a fixed* $n > 2$ *and for all* $(x_1 ,..., x_n) \in \Omega_n{}^\circ$; $(p_1 ,..., p_n)$, $(q_1 ,..., q_n) \in \Gamma_n{}^\circ$ *if, and only if, there exists a constant* $\alpha \geqslant 0$ *and functions* $\gamma_k \colon B\backslash\{0\} \to R$ $(k = 1,..., n)$ *such that* (11) *holds for all* $x_k \in B\backslash\{0\}$, $p_k \in]0, 1[$ $(k = 1, 2,..., n)$. *Then the right-hand side of* (3) *is again*

$$\alpha \sum_{k=1}^{n} p_k \log p_k + \sum_{k=1}^{n} p_k \gamma_k(x_k).$$

If f_k may depend, in addition to q_k, on the whole "experiment" (partition) $(x_1 ,..., x_k)$, it follows immediately from Theorem 1 that

$$f_k(x_1 ,..., x_n; q) = \alpha(x_1 ,..., x_n) \log q + \gamma_k(x_1 ,..., x_n).$$

In Theorems 2, 3 we have shown that, in the case where f_k depend only upon x_k, while obviously the γ_k also depend just upon x_k, the function α is *constant*.

We conclude by considering the inset case $f_1 = f_2 = \cdots = f_n = f$.

COROLLARY. *The inequality*

$$\sum_{k=1}^{n} p_k f(x_k, q_k) \leqslant \sum_{k=1}^{n} p_k f(x_k, p_k) \tag{13}$$

holds for a fixed $n > 2$ *and for all* $(x_1, ..., x_n) \in \Omega_n$ *(or* $\Omega_n{}^\circ$*);* $(p_1, ..., p_n)$, $(q_1, ..., q_n) \in \Gamma_n$ *(or* $\Gamma_n{}^\circ$*) if, and only if, there exists a constant* $\alpha \geqslant 0$ *and a function* $\gamma: B \to R$ *such that*

$$\begin{aligned} f(x_k, p_k) = \alpha \log p_k + \gamma(x_k) \qquad & \textit{for all } x_k \in B \ (\textit{or } \in B\backslash\{0\}), p_k \in]0, 1[, \\ & [f(x_k\, 1) \geqslant \gamma(x_k)], \qquad (k = 1, 2, ..., n). \end{aligned} \tag{14}$$

(*with the conventions* (4) *and* $0 \log 0 = 0$). *Thus the right-hand side of* (13) *reduces to*

$$\alpha \sum_{k=1}^{n} p_k \log p_k + \sum_{k=1}^{n} p_k \gamma(x_k). \tag{15}$$

The latter is the sum of the expected value of a random variable and of a constant multiple of the Shannon entropy, just like the inset entropies with certain recursive or branching properties (cf. Aczél and Daróczy, 1978; Ng, 1977).

Forte (1977) has dealt with a somewhat similar but, as he emphasized, different situation where $x_1, x_2, ..., x_n$ are values of a random variable with probabilities $p_1, p_2, ..., p_n$. The present discussion evidently applies to this situation (and more general ones) too and our formula (12) is very similar to formula (40) of Forte (1977). (The difference between (12), (15) and Forte (1977), Ng (1977), Aczél and Daróczy (1978) is the sign of α. This is explained in Aczél and Daróczy (1975, Sect. 4.3).)

ACKNOWLEDGMENTS

This research has been supported in part by the Natural Sciences and Engineering and the Social Sciences and Humanities Research Councils of Canada, and in part by the California Institute of Technology.

REFERENCES

1. J. ACZÉL (1973), On Shannon's inequality, optimal coding, and characterizations of Shannon's and Rényi's entropies (Convegno Informatica Teoretica, Ist. Naz. Alta Mat., Roma 1973), *Symp. Math.* **15** (1975), 153–179.
2. J. ACZÉL (1974), "Keeping the Expert Honest" Revisited—or: A method to prove the differentiability of solutions of functional inequalities, *Selecta Statistica Canadiana* 2, 1–14.

3. J. ACZÉL AND Z. DARÓCZY (1975), "On Measures of Information and Their Characterizations," Academic Press, New York/San Francisco/London.
4. J. ACZÉL AND Z. DARÓCZY (1978), A mixed theory of information, *I*. Symmetric, recursive and measurable entropies of randomized systems of events, *Rev. Française Automat. Inform. Recherche Opérationnelle*; *Informat. Théor.* **12**, 149–153.
5. J. ACZÉL AND A. M. OSTROWSKI (1973), On the characterization of Shannon's entropy by Shannon's inequality, *J. Austral. Math. Soc.* **16**, 368–374.
6. J. ACZÉL AND J. PFANZAGL (1966), Remarks on the measurement of subjective probability and information, *Metrika* **11**, 91–105.
7. P. FISCHER (1972), On the inequality $\sum p_i f(p_i) \geqslant \sum p_i f(q_i)$, *Metrika* **18**, 199–208.
8. B. FORTE (1977), Subadditive entropies for a random variable, *Boll. Un. Mat. Ital. B.* (5) **14**, 118–133.
9. I. J. GOOD (1952), Rational decisions, *J. Roy. Statist. Soc. Ser.* B **14**, 107–114.
10. I. J. GOOD (1954), "Uncertainty and Business Decisions," 2nd ed., Liverpool Univ. Press, Liverpool, 1957.
11. P. R. HALMOS (1950), "Measure Theory," Van Nostrand, Princeton, N.J.
12. J. MARSCHAK (1959), "Remarks on the Economy of Information," (pp. 79–98 (Contrib. Sci. Res. Management, UCLA 1959), Univ. of California Press, Berkeley, 1960.
13. J. MCCARTHY (1956), Measures of the value of information, *Proc. Nat. Acad. Sci. U.S.A.* **42**, 654–655.
14. C. T. NG (1977), Universal parallel composition laws and their representations, *Math. Scand.* **40**, 25–45.
15. W. WALTER (1976), Remark on a paper by Aczél and Ostrowski, *J. Austral. Math. Soc. A.* **22**, 165–166.